图书情报工作
第65卷 第15期 2021年8月

社交媒体用户同意自己的数据被用于研究吗?*
——基于用户接受意向的伦理反思

陈一[1] 李斯[2]

[1] 武汉大学信息管理学院 武汉 430072 [2] 北京大学信息管理系 北京 100871

摘 要: [目的/意义]以社交媒体用户为研究对象,调查他们对于自身的社交媒体数据被采集用于研究的接受意向,并探索熟悉度、信任、利他主义对接受意向的影响,剖析知情同意程序的必要性,同时提醒学者进行伦理反思,为建立社交媒体数据研究伦理规范和与用户开展伦理风险沟通提供参考。[方法/过程]建立用户从知情到同意的心理机制模型,采用问卷调查法收集来自新浪微博用户的数据,共获得320份有效问卷,并对模型进行验证。[结果/结论]验证出两条有效的心理路径:熟悉-利他主义-接受意向和熟悉-信任-利他主义-接受意向。发现熟悉度和利他主义是影响用户接受意向的重要因素。在社交媒体数据研究中,用户有较强的伦理需求,知情同意程序很有必要。研究者还可以主动表达尊重及认可,以提升用户的利他主义,提升用户的接受意向,规避研究过程中的伦理争议。

关键词: 伦理 社交媒体数据 知情同意 熟悉度 利他主义 信任
分类号: B829
DOI: 10.13266/j.issn.0252-3116.2021.15.014

1 引言

社交媒体平台包含着海量的用户行为数据,越来越多的研究人员采集和利用这类数据来探索用户在虚拟世界中的行为[1]。纽约时报曾发文表示,学者们对Facebook、谷歌等企业收集的大量个人数据的使用前景感到兴奋,他们认为这些数据可能会改变社会科学研究[2]。当前,包括社交媒体数据在内的网络数据研究已成为重要的研究数据来源,这种新型的定量和统计分析已经成为最普遍的社会科学研究形式之一[3]。

用户的社交媒体账户中包含着大量的个人信息。在Twitter中,40%-50%的推文中包含有关作者的信息[4],其中还有可能包含其联系人的相关数据,如个人身份信息、位置数据、健康信息以及图像等。有研究显示,我国新浪微博用户基本注册信息真实程度最高为100%,最低为17.3%[5],即便被采集的数据不包含个人身份信息,但通过关联分析,掌握该用户较全面的个人基本信息甚至个人私密信息并不困难[6-7]。

传统的人类受试者研究需要包括知情同意等程序,以确保"参与者了解他们在做什么、会有什么局限、会带来什么风险"[8],是科学研究的关键环节。但互联网的发展为科学研究引入了一系列额外的复杂问题,如知情同意[9]、服务条款[10-11]、与参与者之间的关系[12]、公共空间和公共关系的界定[13]等。在使用社交媒体数据开展的研究中,是否需要知情同意程序还存在争议,数据属性的模糊性以及获取的数据的便捷性使得科研人员往往忽略了知情同意程序,并且他们也并不认为其中可能存在侵犯隐私、违背伦理规范等问题[14-15]。因而,许多用户往往在不知情的情况下就被动参与到研究中。在实践中,有些研究已经受到了社交媒体用户的抗议,他们认为这类研究不符合伦理规范,侵犯了个人隐私[16-17]。

虽然新的研究方法和数据收集方法不断发展,但在线研究的伦理准则和政策制定仍然处于初级阶段[18]。尽管一些国际组织已经为研究人员提供了相关指南[19],但是这些指南的制定却很少参考用户的意

* 本文系科技大数据湖北省重点实验室(中国科学院武汉文献情报中心)开放基金"基于机器学习的未来导向技术预见研究"(项目编号:20KF011006)研究成果之一。
作者简介: 陈一(ORCID:0000-0003-1828-1165),副教授,博士后;李斯(ORCID:0000-0002-4627-9556),博士后,助理研究员,通讯作者,E-mail:pkulisi@pku.edu.cn。
收稿日期:2020-12-22 修回日期:2021-05-08 本文起止页码:120-129 本文责任编辑:徐健

见[20],且较少有规范的知情同意程序,为研究提供伦理指导。有学者表示,基于社交媒体数据的研究在不断增长,迫切需要提高研究者对其伦理挑战的认识[21]。B. Brown 等学者认为,研究伦理应该建立在"被研究者的敏感性"和"日常实践"的基础上,而不仅仅是法律问题[22]。鉴于此,调查社交媒体用户对于此类研究的接受意向,并探索影响因素,应该成为制定伦理准则和规范此类研究的基础。

新浪微博是全球最具影响力的中文社交媒体,截至 2020 年 11 月,月活跃用户达 5.23 亿,日活跃用户量达 2.29 亿。笔者将新浪微博用户为对象,调查他们对于自身社交媒体数据被采集并应用于研究的真实看法,探索影响用户对社交媒体数据研究接受意向的因素,以提醒学者在从事类似研究时进行伦理反思,也希望能够为建立社交媒体数据研究伦理规范和与用户开展伦理风险沟通提供参考。

2 研究综述

目前,未经用户同意的"社交媒体数据研究"是否合乎伦理正在引发一场持续性的争辩[23]。一些学者认为知情同意程序没有必要,因为社交媒体中的数据是公开的,用户有意识地选择让全世界看到他的活动,如果可以不受限的访问某个人的账户,这就意味着该用户没有隐私期待[24]。但有研究表明,用户也并不像研究者设想的这样缺乏伦理期待。有学者未经用户允许使用 Facebook 用户数据进行"情绪感染"实验,虽然 Facebook 的研究涉及直接实验而不是收集现有数据,但是也引发了公众的强烈抗议[8,25]。C. Fiesler 和 N. Proferes 的调查显示,有超过一半的受访者不接受采集自己的社交媒体数据用于研究,或者表示会视情况而定[20]。M. Zimmer 发现,尽管研究人员认为 Twitter 是公开的,但用户同意公众查看推文并不意味着同意收集和分析[25]。D. Nunan 和 B. Yenicioglu 认为目前大多社交媒体数据研究是"不知情同意",这造成了伦理模糊,他们认为"知情同意"并不是一种限制,而是能够让受访者充分参与研究过程,并免受潜在伤害[26]。在实践中也还没有形成成熟的知情同意程序。B. Custers 等表示现行的很多法规都涉及同意的许多方面,但大多只规定了非常笼统的同意范围,并没有关注同意程序的具体细节,在某些情况下,抽象的法律法规与社交媒体框架和隐私声明的具体实施之间存在着脱节[27]。

目前也有一些研究探索了影响用户接受这类研究的因素。包括人口学统计因素、用户认知因素等。M. L. Williams 等研究了参与者的人口统计学因素与他们接受数据用于科学研究的关系,包括性别、年龄、种族、收入情况、性取向等[28]。关于认知因素的研究主要探讨了参与者对社交媒体数据用于科学研究的了解程度,多数研究认为社交媒体用户对社交媒体平台并不熟悉,更不知道自己的数据可能会被用于科学研究,因此当得知自身数据被采集利用时会感到非常反感。N. Proferes 的研究表明,用户普遍不知道 API 的存在,甚至一些用户可能没有意识到自己的推文是完全公开可见的[29]。Twitter 用户也常常拥有自己想象不到的"粉丝",当用户在平台上发布个人消息时,这些意想不到的"粉丝"可能会带来一定风险[30]。另外,S. S. J. Lee 认为解决使用社交媒体数据研究的伦理问题与信任相关,他认为研究是基于关系的,维护关系的一个核心挑战就是解决信任问题[31]。也有一些研究分析了影响用户主动分享信息的相关因素,如熟悉度的增加会提高用户参与某项目的意愿[32-33],利他主义价值观会增加人们在网络社区中分享知识的意愿[34-35]。

综上所述,已经有不少研究针对社交媒体数据研究中伦理问题展开探讨,然而,作为社交媒体数据研究最直接的利益相关者,用户大多对于自身数据被采集使用并不知情,也对未经允许就采集数据的行为表示反对。当前有关用户对社交媒体数据研究的接受意向以及影响因素的探索都还不够深入。在影响社交媒体用户认知因素的研究中,虽然研究者基于自身经验或描述性统计数据提出参与者对于社交媒体数据研究的熟悉程度是重要的因素,信任因素也被研究者提及,但都没有经过实证数据的检验。鉴于此,笔者将调查分析用户是否接受自身社交媒体数据用于研究,并探索哪些因素影响了用户的同意过程,希望为制定科学的数据使用规范和伦理准则提出提供参考,也提醒研究者在新的数据研究环境下进行伦理反思。

3 研究过程

3.1 理论背景

M. Fishbein 和 I. Ajzen 将行为意向定义为个体想要采取某一特定行为的行动倾向[36],它是一种认知活动,反映个人从事某项行为的意愿与有意识的计划[37]。两位学者提出的理性行为理论和计划行为理论强调了认知因素在决定行为意向时的重要性,他们假设人是理性的,会根据已有的信息并通过信息处理来形成行为信念,从而影响行为意向的形成[37-38]。从

理性行为理论和计划行为理论视角看，用户同意将数据用于科学研究可被认为是他们对这类研究的接受意向。因此，笔者假设参与者是理性主体，探讨信息因素和信念因素如何影响他们的接受数据用于科学研究的意向。基于此，笔者建立了用户接受数据用于科学研究的意向的影响因素模型，以用户对社交媒体数据的熟悉度作为信息因素，以用户对社交媒体数据研究的接受意向作为同意程度的测量。在信念因素方面，已有研究证明了利他主义价值观影响人们分享知识的意愿[39]，用户接受将数据用于科学研究也可被视为一种信息分享意愿，因而笔者假设利他主义对人们的接受意向也存在影响。同时，信任因素也被证明是影响人们行为意向的前置因素[40]，因此笔者设定信任和利他主义作为影响接受意向的信念因素。最终模型如图1所示，具体变量间的关系阐述见3.2。

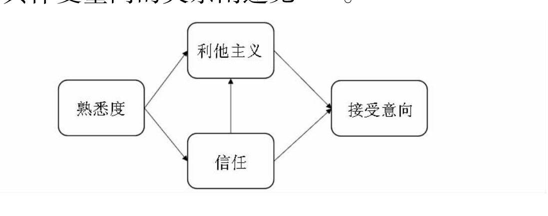


图1 研究假设模型

3.2 研究假设

（1）利他主义影响接受意向。利他主义可能是用户接受社交媒体数据用于科学研究的直接影响因素。在研究志愿者参与科学研究意向的研究中，利他主义被证明是志愿者参与意愿的影响因素之一[33,41]。已有研究表明，利他的人格特质是社交媒体用户在社交媒体平台共享知识的动机[39]。在社交媒体数据用于科学研究情境中，参与者几乎不能获得任何直接的回报，他们可能会出于帮助他人的动机而更容易接受将自己的数据用于科学研究。

基于此，提出假设1：用户的利他主义正向影响其对社交媒体数据研究的接受意向。

（2）对研究者的信任影响接受意向。信任是指基于对人们行动结果的良好认识而对人们开放的行为[42-43]，是减少复杂性的有效办法之一[44]，有了信任，人们往往会做出冒险行为。T. Thornton[45]的研究指出，在公众 科学项目中，信任是实现双方有效合作的前提条件。在社交媒体数据研究中，相比于传统的人类受试者研究，社交媒体用户与研究者处于弱关系中，信任取决于用户对研究者的动机和专业水平的判断。如果用户认为科研人员将来不会以不当的方式使用社交媒体数据，即对研究者有较高的信任程度，那么他们可能接受社交媒体数据研究。

因此，提出假设2：用户对研究者的信任程度正向显著影响对社交媒体数据研究的接受意向。

（3）熟悉度与信任、利他主义、接受意向的关系。熟悉是一种理解，通常是基于先前的互动和经验，以及对他人做什么、为什么做、什么时候做的了解[44]，可以减少不确定性[46]。

熟悉是信任的先决条件[44]，因为熟悉可以通过减少不确定性来增加他人或对象的信任[47]。熟悉会在经验有利时创建信任，或者在不信任时破坏信任[46]。人们参与临床研究时，对研究过程的相对熟悉度会使得对研究人员的信任度更高[47]。用户对社交媒体平台和社交媒体数据研究的熟悉程度是基于先前的经验而产生的。如果用户了解社交媒体平台的各项功能、知晓其限制条款、熟悉科研人员如何采集和使用社交媒体数据，并且这些经验是有利的，就能够降低他们对于自身数据被采集利用的疑虑，基于先前的经验，用户可能更加信任研究者。

因此，提出假设3：熟悉度正向显著影响对科研人员的信任程度。

基于上述对熟悉度与信任、信任与行为意向关系的阐述，可以认为熟悉度可能通过信任的中介作用影响接受意向。熟悉涉及对他人当前行为的理解，而信任则涉及对他人未来行为的信念[44]。如果用户熟悉科研人员如何采集和使用社交媒体数据，了解他们出于何种目的，就能够降低他们对于自身数据被采集利用的疑虑，因此可能更加信任研究者，从而提升他们对社交媒体数据用于科学研究的接受意向。

假设4：熟悉度通过信任的中介作用正向显著影响用户对社交媒体数据研究的接受意向。

熟悉度还会影响用户的利他主义价值观。有学者指出，熟悉度是利他主义的关键影响因素[48-49]，熟悉感可以增加人们的利他主义[50]。在本研究情境中，用户对社交媒体数据研究的熟悉程度越高，就越了解这一行为对于他人和社会的重要意义，这种对社交媒体数据研究积极后果的认知与用户的利他主义价值观相契合，更加增强了用户的利他主义价值观，使得他们对研究中所产生的伦理问题有较高的接受意向。

基于此，本研究提出假设5：熟悉度正向显著影响用户的利他主义。

假设6：熟悉度通过利他主义的中介作用正向显著影响用户对社交媒体数据研究的接受意向。

熟悉度对接受意向的影响还可能是由信任和利他

所介导的。首先,假设3论证了熟悉度对信任产生正向影响。此外,也有研究发现,人们的信任水平正向显著影响利他主义[51]。社区信任通过利他主义的中介作用影响用户在社交媒体中上的知识共享意愿。当个人在虚拟社区中感受到信任的气氛时,利他主义水平高的人比利他主义水平低的人更有可能自由共享信息或讨论社区中的个人经验[52]。

在本研究情境中,如果用户对社交媒体平台及社交媒体数据研究较为熟悉,积极的经验能够提升他们对科研人员的信任程度,进而提升其利他主义,帮助科学研究的意愿得以增强,对社交媒体数据研究接受意愿就越高。因而,从熟悉度到接受意向的路径是合理的。

因此提出假设7:对科研人员的信任正向显著影响利他主义。

假设8:熟悉度对接受意向的正向影响由信任和利他的双重中介,即用户对社交媒体平台和社交媒体数据研究越熟悉,就能够更加信任研究者,利他性越强,对社交媒体数据研究的接受意向越强。

3.3 问卷设计

为保证测量的有效性,本研究的量表均来源于已有的成熟量表或研究成果,并结合本研究的情境进行适当改编,表1呈现了各变量的测量题项及来源。问卷采用李克特5级量表进行调查,其中接受意向的强度衡量了用户对于社交媒体数据研究接受或反对程度,5表示“非常接受”,4表示“接受”…1表示 “非常反对”;信任和利他主义这2个变量请受访者填写与自身情况的符合程度,5表示“非常符合”,4表示“符合”…1表示“非常不符合”;熟悉度这一变量考察用户的熟悉程度,5表示“非常熟悉”,4表示“熟悉”…1表示“非常不熟悉”。

表1 各变量的测量题项及来源

变量	编码	题项	来源文献
信任	T1	我相信研究者不会将我的微博数据用作他途	通过访谈自拟
	T2	我相信研究者在研究完成后能后妥善保管我的微博数据	
	T3	我相信研究者只会采集获得我同意的微博数据	
	T4	我相信研究者如果对我的微博数据有进一步其他使用会再次征求我的同意	
利他	A1	我愿意通过提供微博数据来帮助科学研究	改编自耿瑞利等[53]
	A2	我的微博数据能够帮助到科学研究让我感觉很棒	
	A3	我的微博数据可能会帮助到某项科学研究	
熟悉度	F1	您的微博可能会被用于科学研究	改编自 M. Zimmer & N. J. Proferes[15]; C. Fiesler & N. Proferes[20]
	F2	微博的 API 接口及功能	
	F3	微博使用过程中的一些限制条款	
	F4	没有隐私设置的微博是完全公开的	
接受意向	AI1	您如何看待微博数据被用于科学研究	改编自 C. Fiesler & N. Proferes[20]
	AI2	如果您的某些微博被采集用于科学研究,您的感受是	
	AI3	如果您的所有微博发布历史都被采集用于研究,您的感受是	

3.4 数据收集

研究组于2019年8月13日利用“问卷星”平台发放预调研问卷,根据受访者的反馈意见修改了问卷中的一些题项和模糊概念,以确保正式调查时被调查对象能够很好地理解问卷题项。然后进行了大范围的正式调查。为保证样本的代表性,邀请不同年龄、不同性别、教育层级与职业背景的微博用户填写问卷,以尽可能广泛覆盖调查对象。调查持续了6周,最终回收问卷385份,其中有效问卷320份(具体样本见表2)。样本分布于全国21个省(自治区、直辖市、特别行政区),47个市。受调查者59.4%为女性,40.6%为男性,年龄分布在15－60岁之间,其中21－30年龄段所占比例最大,达到60.3%。有接近60%的受访者在接受调查1个月内发布了微博。有超过75%的受访者注册微博时间达到4年以上,受访者中绝大多数(76%)只有1个微博账户。微博为用户提供了不同的隐私设置选项,如限定“我可以收到哪些人的@”“允许哪些人评论”“微博可见时间范围”等,其中有85名受访者限定了“微博可见时间范围”,所占比例最大,有207位受访者对自己的账户没有进行任何隐私设置。

3.5 信度和效度检验

表3显示了量表的信度和收敛效度检验结果,从中可以看出,各维度的 Cronbach's Alpha 系数均大于0.7,表明量表具有良好的信度。

效度包括内容效度和结构效度。本研究变量的测量题项主要来源于现有的量表或文献,经过严谨翻译和预调研测试,并且经过专家修改,因此具有较好内容效度。结构效度包含收敛效度和区分效度,本研究使用验证性因子分析得出检验两类效度的各项指标。

表 2 调查对象基本信息统计

测量	题项	人数	百分比/%
性别	男	130	40.6
	女	190	59.4
年龄	20 岁以下	67	20.9
	21－30 岁	193	60.3
	31－45 岁	39	12.2
	46 岁以上	21	6.6
学历	初高中	30	9.4
	大专	16	5
	本科	138	43.1
	硕士	96	30
	博士	40	12.5
微博使用频率	每天 5 次以上	65	20.3
	每天 2－4 次	81	25.3
	每天 1 次	34	10.6
	多天 1 次	140	43.8
最近一条微博发布时间	受访当天	23	7.2
	1 周以内	111	34.7
	1 个月以内	49	15.3
	3 个月以内	22	6.9
	半年以内	17	5.3
	半年以上	98	30.6

表 3 量表的信度和收敛效度检验(N＝320)

维度	题项	载荷值	α	CR	AVE
信任	T2	0.810	.923	0.927	0.808
	T3	0.939			
	T4	0.942			
利他	AU1	0.912	.934	0.935	0.828
	AU2	0.913			
	AU3	0.904			
熟悉度	F1	0.699	.817	0.822	0.607
	F2	0.793			
	F3	0.839			
接受意向	AI1	0.856	.868	0.878	0.708
	AI2	0.912			
	AI3	0.747			

在使用验证性因子分析获得检验效度的指标时对测量模型进行了修正[54]。根据修正指引,信任中的 T1 题项和熟悉度中的 F4 题项与其他题项的相关度较大,因此将其删除,再进行验证性因子分析,删除后各项指标得到一定改善,超过或接近推荐值。

收敛效度是指同一个潜变量的各个测量指标间相互关联的程度。检验指标有因子载荷量、平均方差抽取量(AVE)和组合信度(CR)。表 3 显示了各指标的值,从中可以看出,每个维度的组合信度 CR 均大于 0.7,说明各维度中的题项较一致地解释了该潜变量。并且每个维度的 AVE 大于 0.5,表明各维度收敛效度良好。

区分效度的评估指标是潜变量(维度)的 AVE 平方根与该潜变量与其他潜变量的相关系数的比较。表 4 中 AVE 的平方根均大于潜变量间的相关系数,说明本研究构建的量表具有良好的区分效度。

表 4 AVE 平方根与潜在变量间的相关系数(N＝320)

潜变量	1	2	3	4
1 信任	**0.899**			
2 利他	0.597	**0.910**		
3 熟悉度	0.304	0.357	**0.779**	
4 接受意向	0.363	0.570	0.232	**0.841**

在经过信度和效度分析后,可以认为本研究的测量模型和样本数据具有良好的信效度,为接下来的假设检验提供了良好基础。

4 研究结果

4.1 用户对社交媒体数据研究的接受意向

在 320 名受访者中,绝大多数用户对“微博数据用于科学研究”并没有表现出明确的反对,仅有约 10% 的用户表达了反对意见。但针对“如果您的所有微博发布历史都被采集用于研究”这一问题,受访者的态度发生明显变化,其中有 40% 的用户明确表示反对或抗议,27.5% 的用户持中立模糊态度。

在不同的场景下,用户对这类研究的接受程度也存在差异。受访者对于“在没有被告知的情况下,您的微博被采集使用”非常反对的比例达到 42.8%,而如果是“采集使用后告知”,这一比例就下降至 15.3%;相较于被采集的数据是“计算机分析”“人工阅读”受到更大的排斥,选择“非常反对”和“反对”的比例分别都达到了前者的 3 倍左右;针对“标明数据来源”和“匿名处理”这两个问题,受访者明显更倾向于“匿名处理”,不希望识别自己的真实身份。有受访者表示:“如果本人同意可以,未经允许就采集令人反感”“不希望不告知就使用人们的微博数据”。

4.2 用户对社交媒体数据研究接受意向的影响因素模型验证

4.2.1 各变量间的相关性

表 5 显示测量变量之间的相关系数。相关分析的结果表明,所有变量之间的相关关系均显著,信任、利他主义和熟悉度与接受意向均为正相关关系。

陈一，李斯. 社交媒体用户同意自己的数据被用于研究吗？——基于用户接受意向的伦理反思[J]. 图书情报工作,2021,65(15):120–129.

表 5　变量间的相关关系(N=320)

变量	1	2	3	4
1 信任	1			
2 利他	0.597	1		
3 熟悉度	0.304	0.357	1	
4 接受意向	0.363	0.570	0.232	1

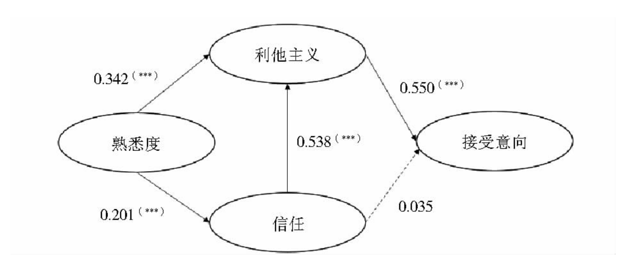


图 2　假设检验结果

4.2.2　直接效应检验

在使用 AMOS 软件进行检验后，各项参数均接近或大于推荐值(x2/df = 2.205；GFI = 0.949；AGFI = 0.919；NFI = 0.962；TLI = 0.979；FMIN = 0.339；RMSEA = 0.061)，说明模型拟合良好。表 6 和图 2 显示了假设检验结果，除了信任对接受意向的影响不显著，其余假设均成立。

表 6　直接效应路径系数(N=320)

假设	作用路径	非标准化系数	p	标准化系数	标准差	T 值
H1	利他主义→接受意向	.522	***	.550	.067	7.831
H2	信任→接受意向	.030	.605	.035	.058	.518
H3	熟悉度→信任	.342	***	.304	.070	4.870
H5	熟悉度→利他主义	.201	***	.195	.057	3.512
H7	信任→利他主义	.494	***	.538	.050	9.954

注：***p <0.001

4.2.3　中介效应检验

由于在直接效应检验中，信任对接受意向的直接影响不显著，因而假设 4 熟悉度通过信任的中介作用影响接受意向不成立，在中介效应检验中将不检验该假设。

本研究采用 Bootstrap 程序就中介效应进行检验。利用重复随机抽样的方法在原始数据(N = 320)中抽取 5 000 个 Bootstrap。在表 7 中，各路径在 95% 置信区间上下限均不包含 0，表明所测量的中介效应皆显著。熟悉度通过利他主义的中介作用影响接受意向(b = –0.121，Boot SE = 0.043，Bias-corrected 95% CI = [0.047，0.214]，Percentile 95% CI = [0.044，0.209])，且间接效应为正，假设 6 被证实。熟悉度还通过信任和利他的中介作用影响接受意向(b = 0.102，Boot SE = 0.029，Bias – corrected 95% CI = [0.058，0.179]，Percentile 95 CI = [0.048，0.163])，且间接效应为正，假设 8 被证实。

表 7　间接效应的 Bootstrap 验证结果(N=320)

假设	路径	效应值	SE	Bias-corrected 95% CI			Percentile 95% CI		
				Lower	Upper	p	Lower	Upper	p
H6	熟悉度→利他→接受意向	0.121	0.043	0.047	0.214	0.001	0.044	0.209	0.001
H8	熟悉度→信任→利他→接受意向	0.102	0.029	0.058	0.179	0.000	0.048	0.163	0.000
总效应	0.235	0.051	0.149	0.353	0.000	0.139	0.340	0.000	

5　讨论及启示

大量的用户行为数据给研究者提供了丰富的研究资源，但大多数时候用户是在不知情的情况下参与到了研究中。笔者探索了用户对社交媒体数据研究的接受意向及其影响因素，通过调查发现，实际上并不是所有社交媒体用户都同意并接受这种研究行为。研究者们需要更多地考虑公众的需求，在科技创新与伦理规范之间寻求平衡。

5.1　理论贡献

本研究的调查结果显示，绝大多数受访者对于研究者不告知就采集社交媒体数据的行为表示反对或非常反对，他们大多是在不知情、被动的情况下就成为研究"参与者"。这与之前 M. Zimmer 对 Twitter 用户的研究结果一致[25]，即用户在社交媒体上公开发布数据，并不等同于"同意"自己的数据被采集使用。

本研究探索了影响用户接受社交媒体数据研究的因素，发现熟悉度对信任和利他主义都产生了正向显著的影响，这再一次验证了熟悉度是信任的先决条件[52–53]，以及熟悉度为利他主义提供了支持[54]。当用户对平台各种功能和限制较为熟悉，并知晓这类研究的流程和目的时，他们会出于认同科学研究的意义而使利他主义价值观得以增强，从而更愿意接受将数据用于科学研究；在另一条路径中，则需要信任变量的参

与,当用户的熟悉度越高,就越信任研究人员,相信他们会妥善使用和保管数据,帮助科学研究的动机得以增强,从而能接受将自己的社交媒体数据用于科学研究。

在两条被验证的路径中,利他主义均是用户同意将社交媒体数据用于科学研究的影响因素。在已有的关于环境保护、绿色商品购买、器官捐献、参与医学研究、共享医学数据、知识共享等亲社会行为的研究中,利他主义都是重要的变量。例如,利他主义对于环境保护态度和器官捐献态度具有积极的影响[55],对于在社交媒体上共享知识的态度和意愿也有积极的影响[39]。而在关于社交媒体数据用于科学研究的文献中还没有探讨利他主义对于用户接受意向的影响。本研究揭示了利他主义与用户接受意向之间的作用路径,证实了利他主义显著正向影响用户对社交媒体数据研究的接受意向。用户对社交媒体数据研究价值的认同会消减他们对于社交媒体数据研究的抵触情绪,出于表达社会责任感以获得内在的满足等目的,更有可能做出对他人更有利的选择。信任没有直接影响用户的接受意向,但依然通过利他主义的影响接受意向,说明信任在弱关系的研究情境中仍然是较为重要的因素。

5.2 实践启示

本研究调查了用户的接受意向,发现知情同意是必要的;用户在微博上发布数据,并不等同于"同意"自己的数据被采集使用;而熟悉度和利他主义是影响用户接受意向的关键因素。在实践中,研究者及相关组织机构应该积极采取举措来规避研究过程中存在的伦理争议。

通过知情同意程序提高用户的接受意向。研究发现,熟悉度通过信任和利他主义的中介作用影响用户的接受意向。"知情同意"涉及向潜在参与者提供有关研究所有方面的明确解释,使其能够完全理解参与者的角色,信任研究人员[56]。让用户"知情",使得用户对于自身数据被采集利用的不确定性担心变得可预测,是提升用户熟悉度的过程。知情同意程序也是一个建立信任的过程[57]。如前所述,用户同意公众查看推文并不意味着同意收集和分析[20],研究者可通过这一程序向潜在参与者传达信息建立信任关系,提升对研究者的信任。有受访者就表示,"只要被告知,很乐意加入学术研究"。提高熟悉度是增强用户信任和利他主义的有效办法。因此,可以通过知情同意程序,让用户熟悉研究过程、了解潜在不确定性、知悉相关权利,进而建立信任,构建良性规范的研究环境。

社交媒体平台在知情同意程序中也应该发挥积极作用。目前,社交媒体平台发布的相关限制条款,如《个人信息保护政策》《开发者协议》中均没有对学术研究型数据采集有明确规范,这也给实际操作留下了很大的不确定性空间。而在国外社交媒体平台上就有对此类使用目的的明确规定,如 Twitter 的用户隐私政策就明确向用户表示研究者可以使用您的公开数据,利用此数据分析趋势或用户见解,告知用户在公开信息时应谨慎思考[58]。社交媒体应该成为研究者和用户之间有用的信息平台,向研究者和用户双方提供正向引导,制定针对研究目的的数据采集使用条款并主动推送给用户,能够规范研究者们的数据利用行为,降低研究过程中涉及的伦理风险以及事后的伦理争议。另外,也能够降低用户顾虑,增强用户对社交媒体平台和社交媒体数据研究的信任和熟悉度,以引导用户对科学研究的积极态度。

主动表达尊重及认可以提升用户的利他主义。调查结果显示,绝大多数受访者对于研究者不告知就采集社交媒体数据的行为表示反对或非常反对。用户们明确的反对态度也显示出较强烈的伦理需求。研究者不能当然地认为社交媒体数据属于公共空间而"假设用户同意"。而大数据环境下,传统的社会、学术组织和监管部门对公私差距的划分可能已经不适用于当下网络环境。实际上,社交媒体数据是公开的还是私人的在某种程度上取决于在线设置本身,以及社交媒体用户对隐私是否有合理的期待[26]。而本研究发现利他主义是影响用户接受意向的主要因素,利他主义强的用户对社交媒体数据研究表现出较高的接受程度。有受访者明确表示,"如果对学术研究有所帮助,提供微博数据是可行的"。研究者可以通过提升用户利他主义价值观来预防研究过程中的伦理争议。R. R. Cottone 和 R. E. Claus[59] 指出,人的伦理态度除了受到自己个人道德原则的影响外,还受到"外部因素"的影响,即他与社会的互动,如利益相关者的要求、国家法律、社会期望和共识、个人或群体的得失等。研究者可以通过调整"外部因素",来满足用户的伦理期待。研究者在数据采集阶段明确表达对用户的尊重并肯定其奉献精神,增进用户的利他主义价值观,进而增强其对社交媒体数据研究的接受意向。

6 结语

知情同意是指应当在充分告知数据主体并征得其

同意的前提下，再对个人数据进行收集、处理、利用和传输等操作。而在大数据环境下，庞大的数据体量以及模糊的数据属性，使得知情同意的有效性、必要性、真实性和经济性也受到质疑，但这并不意味着研究人员可以无视研究伦理，随意采集和利用用户的网络数据。本研究并不是要建议禁止这种研究方式来解决这一困境，完全不允许在研究中使用网络平台数据和没有任何伦理考量地使用网络平台数据可能同样会带来糟糕的后果[20]。伦理决策是一个审议过程，研究人员应该在这个过程中尽可能多地将各种参与人员和资源考虑在内，包括研究合作者、参与者、研究评审委员会、伦理指南、适用的法律先例等[19]。本研究作为一项探索性研究，主要从用户视角，探索了影响用户对社交媒体数据接受意向的认知因素。为了让理论更加完善，后续研究会引入更多的认知因素，也会进一步考虑到外部环境因素的影响，为下一步探索科学的知情同意程序和制定相关伦理规范提供参考。

参考文献：

[1] JAMES N, BUSHER H. Ethical issues in online research[J]. Educational research & evaluation, 2015,21(2):89－94.

[2] GOEL V. As data overflows online, researchersgrapple with ethics [N]. The New York Times, 2014－08－12(1).

[3] KOENE A, PEREZ E, CARTER C J, et al. Research ethics and public trust, preconditions for continued growth of internet mediated research: public confidence in internet mediate research[C]// International Conference on Information Systems Security and Privacy. Piscataway: IEEE, 2016:163－168.

[4] NAAMAN M, BOASE J, LAI C H. Is it really about me? Message content in social awareness streams[C]// Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. New York: ACM, 2010.

[5] 常文英, 刘冰. 基于可信度分析的微博用户个人信息泄露实证研究——以新浪微博为例[J]. 情报杂志, 2015(8):169－176.

[6] CHWARTZ P M, SOLOVE D J. The pii problem: privacy and a new concept of personally identifiable information[J]. Social science electronic publishing, 2011,86(6):1814－1894.

[7] British Psychological Society[EB/OL]. [2021－04－08]. https://www.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research-2017.

[8] DOUGLAS M D. Should internet researchers use Ill-Gotten information? [J]. Science & engineering ethics, 2018, 24:1221－1240.

[9] BAROCAS S, NISSENBAUM H. Big data's end run around anonymity and consent[M]//LANE J, STODDEN V, BENDER S, et al, Eds. Privacy, big data, and the public good: frameworks for engagement. Cambridge: Cambridge Univ ersity Press, 2014: 44－75.

[10] FIESLER C, LAMPE C, BRUCKMAN A S. Reality and perception of copyright terms of service for online content creation[C]// Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW). New York: Association for Computing Machinery,2016: 1450－1461.

[11] VACCARO K, KARAHALIOS K, SANDVING C, et al. Agree or cancel? Research and terms of service compliance [EB/OL]. [2020－01－20]. https://pdfs.semanticscholar.org/3773/15ac09cd7fba7f1f154b0b1481d7c24ad5eb.pdf? _ga = 2.80698775.1012534936.1562862749-228839498.1562862749.

[12] BEAULIEU A, ESTAELLA A. Rethinking research ethics for mediated settings[J]. Information, communication & society, 2012, 15(1):23－42.

[13] ZIMMER M. "But the data is already public": on the ethics of research in facebook[J]. Ethics and information technology, 2010, 12(4):313－325.

[14] VITAK J, PROFERES N, SHILTON K, et al. Ethics regulation inh social computing research: examining the role of institutional review boards[J]. Journal of empirical research on human research ethics, 2017,12(5):372－382.

[15] ZIMMER M, PROFERES N J. A topology of Twitter research: disciplines, methods, and ethics[J]. Aslib journal of information management, 2014, 66(3):250－261.

[16] MCNEAL G S. Facebook manipulated user news feeds to create emotional responses[EB/OL]. [2020－01－20]. https://www.forbes.com/sites/gregorymcneal/2014/06/28/facebook-manipulated-user-news-feeds-to-create-emotional-contagion/#bca47b639dc7.

[17] OKCupid plays with love in user experiments[N]. The New York Times, 2014－07－29.

[18] WARRELL J G, JACOBSEN M. Internet research ethics and the policy gap for ethical practice in online research settings. Canadian journal of higher education, 2014, 44(1): 22－37.

[19] AoIR ethics working committee. Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (Version 2.0) [EB/OL].[2020－11－20]. http://www.aoir.org/reports/ethics.pdf.

[20] FIESLER C, PROFERES N. "Participant" perceptions of Twitter research ethics[J]. Social media society, 2018, 4(1):1－14.

[21] TAYLOR J, PAGLIARI C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? [J]. Research ethics, 2017, 14(2):1－39.

[22] BROWN B, WEILENMANN A, MCMILLAN D, et al. Five provocations for ethical HCI research[C]// Proceeding of the 2016 CHI

Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery,2016:852 – 863.

[23] CLARK K, DUCKHAM M, GUILLEMIN M, et al. Advancing the ethical use of digital data in human research: challenges and strategies to promote ethical practice[J]. Ethics & information technology, 2019,21:59 – 73.

[24] ZIMMER M. Is it ethical to harvest public Twitter accounts without consent? [EB/OL]. [2020 – 11 – 10]. https://www. michaelzimmer. org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/.

[25] KRAMER A D I, GUILLORY J E, HANCOCK J T. Experimental evidence of massive-scale emotional contagion through social networks[J]. Proceedings of the national academy of sciences, 2014, 111(24):8788 – 8790.

[26] NUNAN D, YENICIOGLU B. Informed, uninformed and participative consent in social media research[J]. International journal of market research, 2013, 55(6):791 – 808.

[27] CUSTERS B, SIMONE V D H, SCHERMER B, et al. Informed consent in social media use. The gap between user expectations and EU Personal Data Protection Law[J]. Social science electronic publishing, 2013, 10(4):435 – 457.

[28] WILLIAMS M L, BURNAP P, SLOAN L. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation [J]. Sociology, 2017,51(6):1149 – 1168.

[29] PROFERES N. Information flow solipsism in an exploratory study of beliefs about Twitter. Social media + society [EB/OL]. [2020 – 11 – 10]. https://doi. org/10. 1177/2056305117698493.

[30] LITT E. Knock, Knock. Who's there? The imagined audience [J]. Journal of broadcasting & electronic media, 2012,56(3): 330 – 345.

[31] LEE S S J. Studying "Friends": the ethics of usingsocial media as research platforms[J]. The American journal of bioethics, 2017, 17(3): 1 – 2.

[32] VANUSSA W, JAMES S ASHOK R, et al. Can visual familiarity influence attitudes towards brands? An exploratory study of advergame design and cross-cultural consumer behaviour[J]. Entertainment computing, 2018, 27:194 – 208.

[33] BRANDT D S. "Factors associated with young adults' reported intention of willingness to participate in clinical research"[EB/OL]. [2020 – 04 – 08]. https://ir. uiowa. edu/cgi/viewcontent. cgi? referer = &httpsredir = 1&article = 5604&context = etd.

[34] ZHANG X, LIU S, DENG Z, et al. Knowledge sharing motivations in online health communities: a comparative study of health professionals and normal users [J]. Computers in human behavior, 2017, 75(10):797 – 810.

[35] ACAR O A. Motivations and solution appropriateness in crowdsourcing challenges for innovation [J]. Research policy, 2018, 48(8):103716.

[36] AJZEN I. From intentions to actions: A theory of planned behavior [J]. Advances in experimental social psychology, 1987, 20(8):1 – 63.

[37] FISHBEIN M, AJZEN I. Belief, attitude, intention, and behavior: an introduction to theory and research[M]. Reading, MA: Addison – Wesley, 1975.

[38] ICEK A, DRIVER B L. Application of the theory of planned behavior to leisure choice[J]. Journal of leisure research[J], 1992, 24(3):207 – 224.

[39] OH S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments[J]. Journal of the Association for Information Science and Technology, 2012, 63(3): 543 – 557.

[40] MULLER C E. Examining the inter-relationships between procedural fairness, trust in actors, risk expectations, perceived benefits, and attitudes towards power grid expansion projects[J]. Energy policy, 2020,141:111465.

[41] SENGUPTA S, STRAUSS R P, DEVELLIS R, et al. Factors affecting African-American participation in AIDS research[J]. Jaids journal of acquired immune deficiency syndromes, 2000, 24(3): 275 – 284.

[42] RIEGELSBERGER J, SASSE M A, MCCARTHY J D. Trust at First Sight? A test of users' ability to identify trustworthy e-commerce sites [EB/OL]. [2020 – 04 – 13]. http://sec. cs. ucl. ac. uk/fileadmin/sec/publications/Riegelsberger_Sasse_McCarthy_A_Test_of_Users_Ability_to_Identify_Trustworthy_E-commerce_Sites_HCI2003. pdf.

[43] OKYERE K E, NOR K M. Individual factors and knowledge sharing[J]. American journal of economics and business administration, 2011,3(1): 66 – 72.

[44] GEFEN D. E-commerce: the role of familiarity and trust[J]. Omega, 2000,28(6): 725 – 737.

[45] THORNTON T, LEAHY J. Trust in citizen science research: a case study of the groundwater education through water evaluation & testing program [J]. Jawra journal of the American Water Resources Association, 2012, 48(5):1032 – 1040.

[46] LUHMANN N. Trust and power[M], Chichester, UK: Wiley, 1979.

[47] YOON C, ROLLAND E. Knowledge-sharing in virtual communities: familiarity, anonymity and self-determination theory[J]. Behaviour & information technology, 2012, 31(11): 1133 – 1143.

[48] BODE N W F, MILLER J, O'GORMAN R, et al. Increased costs reduce reciprocal helping behaviour of humans in a virtual evacuation experiment[EB/OL]. [2020 – 05 – 13]. https://www. nature. com/articles/srep15896#citeas.

[49] KHALIL E L. Adam Smith and three theories of altruism[J]. Louvain economic review, 2001, 67(4): 421－435.
[50] KALCHEV D. The impact of positive emotions on altruism in the presence of familiarity[EB/OL]. [2020－05－13]. https://trepo.tuni.fi/bitstream/handle/10024/101633/GRADU-1498464800.pdf? sequence=1.
[51] 朱备. 个体信任水平对利他行为的影响研究[D].金华:浙江师范大学,2016.
[52] CHEN H, FAN H, TSAI C. The Role of community trust and altruism in knowledge sharing: an investigation of a virtual community of teacher professionals[J]. Educational technology & society, 2014,17(3):168－179.
[53] 耿瑞利,申静.不同文化视域下社交网络用户知识共享行为动机研究[J].中国图书馆学报,2019,45(1):60－81.
[54] 钱丽丽. 电子政务公众服务需求及其对系统成功的影响路径研究[D].上海:复旦大学, 2010.
[55] MOSTAFA M M. Altruistic, cognitive and attitudinal determinants of organ donation intention in Egypt: a social marketing perspective[J]. Health marketing quarterly, 2010, 27(1): 97－115.
[56] WILES R, HEATH S, CROW G, et al. Informed consent in social research: a literature review[EB/OL]. [2020－05－05]. http://eprints.ncrm.ac.uk/85/1/MethodsReviewPaperNCRM-001.pdf.
[57] SALMONS J. Getting to yes: informed consent in qualitative social media research[M]//WOODFIELD K, ed. The ethics of online research. Bingley: Emerald Publishing Limited, 2017:109－134.
[58] Twitter privacy policy[EB/OL].[2020－09－20]. https://twitter.com/en/privacy.
[59] COTTONE R R, CLAUS R E. Ethical decision-making models: a review of the literature[J]. Journal of counseling & development, 2000,78(3):275－283.

作者贡献说明：
陈一：确定选题，提出论文整体研究思路与框架，撰写论文；
李斯：数据分析与处理，修改论文。

Do Social Media Users Consent to the Use of Their Social Media Data in Research? ——Ethical Reflection Based on User Acceptance Intention

Chen Yi[1]　Li Si[2]
[1] School of Information Management, Wuhan University, Wuhan 430072
[2] Department of information management, Peking University, Beijing 100871

Abstract: [**Purpose/significance**] This study takes social media users as the research object, investigates their acceptance intention for their own social media data collected for research, explores the influence of familiarity, trust and altruism on acceptance intention, analyzes the necessity of informed consent procedure, and reminds scholars to conduct ethical reflection. It provides a reference for establishing ethical norms for social media data research and conducting ethical risk communication with users. [**Method/process**] The psychological mechanism model of users from informed to consent was established. A total of 320 valid questionnaires were collected from Sina Microblog by questionnaire survey method, and the model was verified. [**Result/conclusion**] This study verifies two effective psychological paths: familiarity-altruism-acceptance intention and familiarity-trust-altruism-acceptance intention. It is found that familiarity and altruism are important factors that affect user acceptance intention. Users have strong ethical needs, and informed consent procedure is necessary. Researchers can also actively express respect and recognition, and enhance users' altruism and acceptance intention to avoid ethical risks in the research process.

Keywords: ethics　social media data　informed consent　familiarity　altruistic　trust